Ann Inst Stat Math (2016) 68:491–539
DOI 10.1007/s10463-015-0503-3

Bootstrapping sample quantiles of discrete data

Carsten Jentsch · Anne Leucht

Received: 13 May 2014 / Revised: 28 October 2014 / Published online: 20 February 2015

Abstract Sample quantiles are consistent estimators for the true quantile and satisfy central limit theorems (CLTs) if the underlying distribution is continuous. If the distribution is discrete, the situation is much more delicate. In this case, sample quantiles are known to be not even consistent in general for the population quantiles. In a motivating example, we show that Efron's bootstrap does not consistently mimic the distribution of sample quantiles even in the discrete independent and identically distributed (i.i.d.) data case. To overcome this bootstrap inconsistency, we provide two different and complementing strategies. In the first part of this paper, we prove that m-out-of-n-type bootstraps do consistently mimic the distribution of sample quantiles in the discrete data case. As the corresponding bootstrap confidence intervals tend to be conservative due to the discreteness of the true distribution, we propose randomization techniques to construct bootstrap confidence sets of asymptotically correct size. In the second part, we consider a continuous modification of the cumulative distribution function and make use of mid-quantiles studied in Ma et al. (Ann Inst Stat Math 63:227–243, 2011). Contrary to ordinary quantiles and due to continuity, mid-quantiles lose their discrete nature and can be estimated consistently. Moreover, Ma et al. (Ann Inst Stat Math 63:227–243, 2011) proved (non-)central limit theorems for i.i.d. data, which we generalize to the time series case. However, as the mid-quantile function fails to be differentiable, classical i.i.d. or block bootstrap methods do not lead to completely satisfactory results and m-out-of-n variants are required here as well. The finite sample

C. Jentsch (✉)
Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany
e-mail: cjentsch@mail.uni-mannheim.de

A. Leucht
Institut für Mathematische Stochastik, Technische Universität Braunschweig, Pockelsstraße 14, 38106 Braunschweig, Germany
e-mail: a.leucht@tu-braunschweig.de

performances of both approaches are illustrated in a simulation study by comparing coverage rates of bootstrap confidence intervals.

Keywords Bootstrap inconsistency · Count processes · Mid-distribution function · m-Out-of-n bootstrap · Integer-valued processes